ORDER RESTRICTED ESTIMATION OF DISTRIBUTIONS WITH CENSORED DATA

Djamel BELLOUT*

The University of Michigan, Ann Arbor, MI 48109, U.S.A.

Abstract: Consider a finite population of hidden objects, and consider searching for them for one unit of time. Suppose that both the size and the discovery time of the objects have unknown distributions, and that the conditional distribution of time given size is exponential with an unknown non-negative and non-decreasing function of the size as the failure rate. Order restricted M.L.E.'s are derived for this function, other parameters are estimated, and the consistency of the estimates is shown.

1. Introduction

Consider searching for hidden objects (like a play of mineral deposits), in a finite population with unknown size N, at a certain cost, and receiving a reward depending on the sizes of the objects found. Let $X_1, X_2, \ldots, X_N$ denote the sizes, and $T_1, T_2, \ldots, T_N$ the discovery times of the objects, in an infinite search. Suppose that (X_1, T_1), $(X_2, T_2), \ldots, (X_N, T_N)$ are independent and identically distributed as (X, T) (non-negative); and let F and G^* denote the unknown distribution of X and conditional distribution of T given X, respectively. Intuitively $G^*(t \mid x)$ should be non-decreasing in x, for each t. That is, it should be easier to find large objects than small ones.

There has been substantial recent interest in this general model. Barouch and Kaufman (1975) described models for exploring petroleum reserves, in which the probability of finding a pool is proportional to its size. Lynden-Bell (1971), and Jackson (1974) derived nonparametric M.L.E.'s for F and G^*, in the case $G^*(t \mid x) = G^*(tx)$. Nicoll and Segal (1980) obtained M.L.E.'s for grouped data, for this model; Bhattacharya, Chernoff and Yang (1983) derived M.L.E.'s for conditional dis-

tributions, based on a conditional likelihood function given the observed x-values, and derived nonparametric estimators of regression parameters in models similar to the one described earlier. Woodroofe (1985) obtained Lynden-Bell nonparametric estimates of F and G from a different perspective and showed their consistency (G being the unconditional distribution of XT in the model above).

Here we study a class of models closely related to those in Kramer (1983), under time censorship. More precisely, $G^*(t \mid x)$ is considered to be of the form:

$$G^*(t \mid x) = 1 - \exp\{-tH(x)\} \tag{1.1}$$

for $t, x > 0$ where H is a positive and non-decreasing function on $(0, \infty)$. Section 2 of this paper describes the M.L.E.'s of H, N and F in the truncated case $T_i \le 1$, $i = 1, 2, \ldots, n$, using the isotonic regression technique. An example is included there. The consistency of the estimates is shown in Section 3.

2. Estimation

As mentioned earlier, this section deals with the derivation of nonparametric M.L.E.'s for F, H, and N in the truncated case $T \le 1$. The primary interest is in a suitable estimation of H. Let n denote the number of i for which $T_i \le 1$; and let $x_1, x_2, \ldots, x_n$ and $t_1, t_2, \ldots, t_n$ denote the sizes of the objects found and their discovery times.

To estimate F, H, and N consider only distribution functions F supported by $x_1, x_2, \ldots, x_n$. And let $p_i = F(x_i) - F(x_i-)$ and $h_i = H(x_i)$, for $i = 1, 2, \ldots, n$. Further let $p = (p_1, p_2, \ldots, p_n)'$, and $h = (h_1, h_2, \ldots, h_n)'$. Then the likelihood function based on this sample is

$$L(p, h, N \mid n; x_1, \ldots, t_n) = \prod_{i=1}^{n} p_i h_i e^{-t_i/h_i}(N)_n (1 - \alpha)^{N-n}, \tag{2.1}$$

where

$$\alpha = \sum_{i=1}^{n} p_i(1 - e^{-h_i}) = P(T \le 1). \tag{2.2}$$

First, we maximize with respect of N, the portion of L that depends on N. Fix values of p and h. Then it is easily seen that the difference $(N)_n(1 - \alpha)^{N-n} - (N-1)_n(1 - \alpha)^{N-1-n}$ vanishes when $N = N_0 = n/\alpha$ and that the maximum occurs when N is an integer adjacent to N_0. The maximizing value is approximated by N_0, for p, and h fixed. Now note that L can be written as

$$L(p, h, N \mid n; x_1, \ldots, t_n) = L_1(p, h \mid n; x_1, \ldots, t_n)L_2(p, h, N \mid n) \tag{2.3}$$

where

$$L_1(p, h \mid n; x_1, \ldots, t_n) = \frac{n!}{\alpha^n} \prod_{i=1}^{n} p_i h_i e^{t_i h_i} \tag{2.4}$$

and

$$L_2(p, h, N \mid n) = \binom{N}{n} \alpha^n (1 - \alpha)^{N-n}. \tag{2.5}$$

Then using Stirling's formula,

$$\max_{N} L_2(\boldsymbol{p}, \boldsymbol{h}, N \mid n) = L_2(\boldsymbol{p}, \boldsymbol{h}, N_0 \mid n)$$

$$\simeq \left[\frac{1}{2\pi n(1-\alpha)} \right]^{1/2}. \tag{2.6}$$

The latter term is not highly sensitive to $\boldsymbol{p}$, and $\boldsymbol{h}$, when compared to L_1. The M.L.E.'s of $\boldsymbol{p}$ and $\boldsymbol{h}$ are obtained from L_1 alone, a further approximation. To find them, fix an $\boldsymbol{h}$ for which $h_i > 0$, $i = 1, 2, \ldots, n$. Then

$$\frac{\partial}{\partial p_i} \log L_1 = \frac{-n}{\alpha} (1 - e^{-h_i}) + \frac{1}{p_i}, \quad i = 1, 2, \ldots, n, \tag{2.7}$$

and $\sum_{i=1}^{n} p_i = 1$. Hence for some Lagrange multiplier λ,

$$\lambda = \frac{1}{p_i} - \frac{n}{\alpha} (1 - e^{-h_i}), \quad i = 1, 2, \ldots, n. \tag{2.8}$$

Multiplying equation (2.8) by p_i and summing over i yields

$$\lambda = n - \frac{n}{\alpha} \sum_{i=1}^{n} p_i(1 - e^{-h_i}) = n - n = 0. \tag{2.9}$$

Hence, the M.L.E. of p_i is

$$\hat{p}_i = \frac{\alpha}{n(1 - e^{-h_i})}, \quad i = 1, 2, \ldots, n, \tag{2.10}$$

and

$$\max_{p} L_1 \propto \prod_{i=1}^{n} \frac{h_i e^{-t_i h_i}}{1 - e^{-h_i}}. \tag{2.11}$$

All that remains now is the M.L.E. of $\boldsymbol{h}$. Now,

$$\frac{\partial}{\partial h_i} \log L_1(\hat{p}, \boldsymbol{h}) = \frac{1}{h_i} - \frac{e^{-h_i}}{1 - e^{-h_i}} - t_i, \quad i = 1, 2, \ldots, n. \tag{2.12}$$

Let

$$u(h) = \frac{1}{h} - \frac{e^{-h}}{1 - e^{-h}}, \quad h > 0. \tag{2.13}$$

Then, u is a decreasing function, as is shown below; so, setting the partial derivatives equal to zero leads to the estimates

$$\hat{h}_i' = u^{-1}(t_i), \quad i = 1, 2, \ldots, n. \tag{2.14}$$

But these estimates are unsatisfactory since they ignore the monotonicity of H. This problem is overcome by the isotonic regression method.

Let us relabel the sample so that $x_1 < x_2 < \cdots < x_n$. Then, $0 \le h_1 \le h_2 \le \cdots \le h_n < \infty$ and $0 \le t_i \le 1$, $i = 1, 2, \ldots, n$. Let

$$\omega_i = -h_i, \quad i = 1, 2, \ldots, n, \tag{2.15}$$

so $-\infty < \omega_n \le \cdots \le \omega_2 \le \omega_1 \le 0$. Let

$$g_1(t \mid \omega_i) = \frac{\omega_i e^{t\omega_i}}{e^{\omega_i} - 1}, \quad i = 1, 2, \ldots, n, \ 0 \le t \le 1, \tag{2.16}$$

and

$$\theta_i = E_{g_1}(T \mid \omega_i) = \frac{e^{\omega_i}}{e^{\omega_i} - 1} - \frac{1}{\omega_i}, \quad i = 1, 2, \ldots, n. \tag{2.17}$$

Then $\frac{1}{2} \ge \theta_1 \ge \theta_2 \ge \cdots \ge \theta_n > 0$; and $\theta_i = u(h_i)$ is decreasing. Hence it suffices to find the M.L.E.'s of the θ_i's. These are based on the unconstrained M.L.E.'s of θ_i, t_i, $i = 1, 2, \ldots, n$. On the other hand, $g_1(t \mid \omega)$ is of the exponential form given in Barlow et al. (1972) p. 92, with

$$\Phi(\theta) = -\left[\log \frac{e^\omega - 1}{\omega} - \omega\theta \right], \tag{2.18}$$

and

$$\phi(\theta) = \Phi'(\theta) = \omega,$$

where θ and ω are related as in (2.17) and ω is regarded as a function of θ. All the conditions imposed by Barlow et al. are satisfied. So, the M.L.E.'s may be determined from Theorem 2.12 of the same book. The order restricted M.L.E.'s of θ_i, which are denoted $\hat{\theta}_i$, are the isotonic regression of t_i, $i = 1, 2, \ldots, n$, with weights $m_i = 1$, $i = 1, 2, \ldots, n$, and are given by

$$\hat{\theta}_i = \min_{s \le i} \max_{r \ge i} \text{Av}(s, r), \quad i = 1, 2, \ldots, n, \tag{2.19}$$

where

$$\text{Av}(s, r) = \sum_{k=s}^{r} t_k / (r - s + 1). \tag{2.20}$$

It is easily seen that $0 \le \theta_i \le \frac{1}{2}$ for all $i = 1, 2, \ldots, n$ and that $\theta_i = \frac{1}{2}$ iff $\omega_i = 0$. To keep the estimators in the parameter space and to avoid problems with the end points, it is convenient to truncate them. Let c_n, $n \ge 1$, be a sequence for which $0 < c_n < \frac{1}{2}$ for all n and $c_n \to 0$ as $n \to \infty$. Then the $\hat{\theta}_i$ of (2.19) may be replaced by $\hat{\theta}_i \wedge (\frac{1}{2} - c_n)$. Thus the estimators of $\theta_1, \theta_2, \ldots, \theta_n$ are taken to be

$$\hat{\theta}_i = \min\{\tfrac{1}{2} - c_n; \min_{s \le i} \max_{r \ge i} \text{Av}(s, r)\}, \quad i = 1, 2, \ldots, n. \tag{2.21}$$

After this, it remains to invert u to obtain the order restricted M.L.E.'s of the h_i's,

$$\hat{h}_i = u^{-1}(\hat{\theta}_i), \quad i = 1, 2, \ldots, n. \tag{2.22}$$

Once h has been estimated, the M.L.E. of p is given by (2.10). We may substitute the estimate of h in (2.10.) and estimate p by

$$\hat{p}_i = \frac{1}{[nC(1 - e^{-\hat{h}_i})]}, \quad i = 1, 2, \ldots, n, \tag{2.23}$$

where $C = \sum_{i=1}^{n} 1/[n(1 - e^{-\hat{h}_i})]$ is a normalizing constant.

Finally the estimate of N is

$$\hat{N} = N_0(\hat{\alpha}_n) = \frac{n}{\hat{\alpha}_n}, \tag{2.24}$$

where $\hat{\alpha}_n = \sum_{i=1}^{n} \hat{p}_i(1 - e^{-\hat{h}_i}) = 1/C$, denotes the M.L.E. of α.

Note. To estimate F in the truncated case $T \le 1$, $F^{\#}$, the conditional distribution of X given $T \le 1$ is estimated by $F_n^{\#}$, the empirical distribution of $x_1, x_2, \ldots, x_n$. But, $dF^{\#}(x) \propto [1 - e^{-H(x)}]dF(x)$. Hence

$$dF_n^{\#}(x) \propto [1 - e^{-\hat{H}_n(x)}]d\hat{F}_n(x) \tag{2.25}$$

where $\hat{H}_n(x_i) = \hat{h}_i$, $i = 1, 2, \ldots, n$, and $\hat{F}_n$ has jump at x_i given by equation (2.23) above.

Example. *The Rimbey–Meadowbrook reef play.* As an illustration, the estimators are applied to estimate the total remaining reserve of an oil play. The data of this example come from the Rimbey–Meadowbrook reef chain located in central Alberta, Canada. This play has been investigated by many workers who proposed models to describe the play, estimate parameters like the population size N, and the size distribution F, and assess the remaining number of undiscovered pools and their potential. They also examine the influence of size on the discovery sequence. Lee and Wang (1986), and Nair and Wang (1987) are good recent sources for such details, and may be consulted for further references. Here x is in millions of barrels.

Assuming that the search started in 1946 and ended in 1970, it lasted 9125 days. The time data were recorded in number of days to go along with algorithm (2.21). For computational convenience the time is reduced to the unit scale. The estimates are reported below and in Table 1. Here $c_n = 0.05$.

$$\hat{\alpha}_n = 0.6715, \qquad \hat{N} = \frac{n}{\hat{\alpha}_n} = \frac{23}{0.6715} \simeq 35.$$

The total remaining reserve is estimated to be 67.96 million barrels. This estimate is substantially higher than that of Nair and Wang (1987) which is 6.11 million barrels. On the other hand it is less than the estimates of Lee and Wang (1986). Our purpose here, however, is more to illustrate the nature of the estimators than to enter a controversy about the oil play.

3. Consistency

Here the consistency of $\hat{H}_n$ is established as a corollary of the consistency of the empirical distributions. The properties of the least concave majorant (L.C.M.) are

Table 1
M.L.E.'s of θ, H and F

i	t_i	x_i	$\hat{\theta}_i$	$\hat{\omega}_i$	$\hat{h}_i$	$\hat{p}_i$	$\hat{F}_n(x)$
1	0.8617	0.3	0.4500	−0.6036	0.6036	0.06443	0.06443
2	0.3859	0.6	0.4500	−0.6036	0.6036	0.06443	0.12896
3	0.7302	1.0	0.4500	−0.6036	0.6036	0.06443	0.19330
4	0.7371	2.2	0.4500	−0.6036	0.6036	0.06443	0.25772
5	0.2059	2.6	0.4500	−0.6036	0.6036	0.06443	0.32215
6	0.7107	3.4	0.4500	−0.6036	0.6036	0.06443	0.38658
7	0.3773	4.7	0.4290	−0.8620	0.8620	0.05054	0.43713
8	0.2860	6.6	0.4290	−0.8620	0.8620	0.05054	0.48767
9	0.2733	8.6	0.4290	−0.8620	0.8620	0.05054	0.53822
10	0.6314	14.2	0.4290	−0.8620	0.8620	0.05054	0.58876
11	0.5771	14.6	0.4290	−0.8620	0.8620	0.05054	0.63930
12	0.2795	14.7	0.3088	−2.5286	2.5286	0.03173	0.67103
13	0.2252	15.0	0.3088	−2.5286	2.5286	0.03173	0.70276
14	0.4216	20.1	0.3088	−2.5286	2.5286	0.03173	0.73450
15	0.2228	27.6	0.2517	−3.5577	3.5577	0.03005	0.76455
16	0.2806	111.0	0.2517	−3.5577	3.5577	0.03005	0.79461
17	0.1810	142.9	0.2175	−4.3322	4.3322	0.02958	0.82420
18	0.2539	169.6	0.2175	−4.3322	4.3322	0.02958	0.85378
19	0.1473	290.8	0.1571	−6.2893	6.2893	0.02925	0.88303
20	0.0447	351.5	0.1571	−6.2893	6.2893	0.02925	0.91230
21	0.2080	366.7	0.1571	−6.2893	6.2893	0.02925	0.94154
22	0.2284	765.9	0.1571	−6.2893	6.2893	0.02925	0.97080
23	0.1023	1295.4	0.1023	−9.7586	9.7586	0.02920	1.00000

also useful. As in the estimation, the main focus here is on the consistency of

$$\hat{H}_n(x_i) = \hat{h}_i = u^{-1}(\hat{\theta}_i), \quad i = 1, 2, \ldots, n, \tag{3.1}$$

where u, $\hat{h}_i$ and $\hat{\theta}_i$ are as in Section 2. Note that the consistency of $\hat{\theta}_i$ follows from that of $\hat{\theta}_i$ since $c_n \to 0$ as $n \to \infty$.

Theorem 3.1. *If the distribution F of X is continuous, and strictly increasing on its support, and H is continuous, then the estimate of H, $\hat{H}_n$, is uniformly consistent. That is*

$$\max_{1 \leq i \leq n} |\hat{H}_n(x_i) - H(x_i)| \to 0 \quad a.s. \ as \ n \to \infty \tag{3.2}$$

or equivalently

$$\max_{1 \leq i \leq n} |\hat{\theta}_i - \theta_i| \to 0 \quad a.s. \ as \ n \to \infty \tag{3.3}$$

where $\theta_i = u(H(x_i))$.

Three lemmas are needed in the proof:

Lemma 3.1 (A.W. Marshall). *Let Γ be concave on $[0, 1]$ and Ψ be a continuous real valued function on $[0, 1]$. If Λ is the L.C.M. of Ψ then*

$$\sup_{0 \leq x \leq 1} |\Lambda(x) - \Gamma(x)| \leq \sup_{0 \leq x \leq 1} |\Psi(x) - \Gamma(x)|. \tag{3.4}$$

For a proof of this lemma see Marshall (1970).

Lemma 3.2. *If Q_n, and Q are increasing functions such that $Q_n \to Q$ and Q is continuous, then Q_n converges to Q uniformly on compact subintervals of the domain.*

This lemma is a version of Polyá's theorem. See Breiman (1968), p. 160.

Lemma 3.3. *Let G, G_n, $n \geq 1$, be concave functions such that $G_n \to G$. Let G'_n be the right hand derivative of G_n, and suppose that G has a continuous derivative G'. Then $G'_n \to G'$, uniformly on any compact subinterval.*

Lemma 3.3 follows from the fact that the derivatives of concave functions converge whenever the functions do, and from Lemma 3.2, which supplies the uniformity.

Let us now introduce the notation needed in the proof of (3.3). Let $x_1, x_2, \ldots, x_n$ be the values of $X_1, X_2, \ldots, X_N$ for which $T_i \leq 1$ (i.e. $n = n_N = \#\{i \leq N : T_i \leq 1\}$), so labelled that $x_1 < x_2 < \cdots < x_n$. Recall that $\theta(x) = E(t_1 \mid x)$. Let $F^\#$ be the conditional distribution of X given $T \leq 1$, and $F_n^\#$ be the empirical distribution of $x_1, x_2, \ldots, x_n$. Let

$$K(y) = \int_{\{F^\#(x) \leq y\}} \theta(x) dF^\#(x), \quad y \in [0, 1]; \tag{3.5}$$

$$K_n(j/n) = \begin{cases} n^{-1} \sum_{i=1}^{j} \theta(x_i) & \text{for } 1 \leq j \leq n, \\ \text{linear} & \text{on } [j/n, (j+1)/n), \ 1 \leq j \leq n-1; \end{cases} \tag{3.6}$$

$$\hat{K}_n(j/n) = \begin{cases} n^{-1} \sum_{i=1}^{j} t_i & \text{for } 1 \leq j \leq n, \\ \text{linear} & \text{on } [j/n, (j+1)/n), \ 1 \leq j \leq n-1, \end{cases} \tag{3.7}$$

and

$$\hat{\hat{K}}_n(j/n) = \begin{cases} n^{-1} \sum_{i=1}^{j} \hat{\theta}_i & \text{for } 1 \leq j \leq n, \\ \text{linear} & \text{on } [j/n, (j+1)/n), \ 1 \leq j \leq n-1. \end{cases} \tag{3.8}$$

Thus $\hat{\hat{K}}_n$ is the L.C.M. of $\hat{K}_n$. It is easily seen that
   (1) K, K_n, and $\hat{\hat{K}}_n$ are concave;
   (2) K, K_n, $\hat{K}_n$, and $\hat{\hat{K}}_n$ are all increasing;
   (3) by consistency of $F_n^\#$,

$$K_n(y) \to K(y) = \int_{\{F^\#(x) \leq y\}} \theta(x) dF^\#(x) \quad \text{w.p. 1,} \tag{3.9}$$

for all y; and

(4) since K_n, K are increasing and K is continuous the convergence in (3.9) is uniform by Lemma 3.2.

Proof of Theorem 3.1. It is enough to prove the result for $\hat{\theta}_i$. That follows by first proving

$$\max_{1 \le j \le n} |\hat{\hat{K}}_n(j/n) - K_n(j/n)| \to 0 \quad \text{a.s.,} \tag{3.10}$$

then by using Lemma 3.3. To prove (3.10), first observe that by Marshall's lemma (Lemma 3.1),

$$\max_{1 \le j \le n} |\hat{\hat{K}}_n(j/n) - K_n(j/n)| \le \max_{1 \le j \le n} |\hat{K}_n(j/n) - K_n(j/n)| \tag{3.11}$$

(the maximum over all y is attained at one of the j/n). Therefore it is enough to show $\max |\hat{K}_n - K_n| \to 0$. Let $\hat{S}_n(y) = n\hat{K}_n(y)$, and $S_n(y) = nK_n(y)$. Then

$$\hat{S}_n(j/n) - S_n(j/n) = \sum_{i=1}^{j} (t_i - \theta_i) \tag{3.12}$$

is the sum of conditionally independent, zero mean random variables given $x_1, x_2, \ldots, x_n$ for each n and j. It follows easily that $\hat{S}_n(j/n) - S_n(j/n)$ is a martingale in j for each n and that

$$E\left[\left(\frac{\hat{S}_n(j/n) - S_n(j/n)}{n} \right)^4 \right] \le \frac{c}{n^2}, \quad 1 \le j \le n, \tag{3.13}$$

for some constant c. So fixing n, and letting $\varepsilon_n = n^{-1/5}$,

$$P\left\{ \max_{1 \le j \le n} |\hat{K}_n(j/n) - K_n(j/n)| \ge \varepsilon_n \right\}$$

$$= P\left\{ \max_{1 \le j \le n} |\hat{S}_n(j/n) - S_n(j/n)| \ge \varepsilon_n n \right\}$$

$$\le \left(\frac{1}{n\varepsilon_n} \right)^4 E[(\hat{S}_n(n/n) - S_n(n/n))^4] \quad \text{(submartingale inequality)}$$

$$\le \frac{c}{n^2 \varepsilon_n^4} \quad \text{(by equation (3.13))}$$

$$= cn^{-6/5} \tag{3.14}$$

and

$$\sum_{n=1}^{\infty} P\left\{ \max_{1 \le j \le n} |\hat{K}_n(j/n) - K_n(j/n)| \ge n^{-1/5} \right\} < \infty. \tag{3.15}$$

Hence by the Borel–Cantelli lemma,

$$P\left\{ \max_{1 \le j \le n} |\hat{K}_n(j/n) - K_n(j/n)| \ge n^{-1/5} \quad \text{i.o.} \right\} = 0. \tag{3.16}$$

This proves (3.10) above. Therefore $\hat{K}_n \to K$ uniformly since, as seen earlier $K_n \to K$ uniformly. Now $\hat{K}_n$ and K are concave, and K is continuously differentiable on the support of F, by the conditions imposed on F and H. So $\hat{K}'_n \to K'$ uniformly by Lemma 3.3, i.e.

$$\max_{y \in [0, 1]} |\hat{K}'_n(y) - K'(y)| \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.17}$$

Similarily,

$$\max_{y \in [0, 1]} |K'_n(y) - K'(y)| \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.18}$$

Equations (3.17) and (3.18) together imply

$$\max_{y \in [0, 1]} |\hat{K}'_n(y) - K'_n(y)| \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.19}$$

Now just notice that $\hat{\theta}_i = \hat{K}'_n(i/n)$, and $\theta_i = K'_n(i/n)$, to conclude that

$$\max_{1 \le i \le n} |\hat{\theta}_i - \theta_i| \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.20}$$

This completes the proof of Theorem 3.1.

In what follows, $\hat{H}_n$ is extended to all of $[0, \infty)$ by letting $\hat{H}_n$ be linear between x_{i-1} and x_i for $i = 1, 2, \ldots, n$ and constant to the right of x_n; $x_0 = 0$.

Corollary 3.1. *If H is continuous and bounded and the support of F is $[0, \infty)$, then*

$$\sup_{0 \le y < \infty} |\hat{H}_n(y) - H(y)| \to 0 \quad \text{a.s. as } n \to \infty. \tag{3.21}$$

Proof. If $0 \le y \le M$, and if n is so large that $x_n > M$, there is an $i = i_n$ for which $x_i \le y < x_{i+1}$ and, therefore, $H(x_i) \le H(y) \le H(x_{i+1})$ and $\hat{H}_n(x_i) \le \hat{H}_n(y) \le \hat{H}_n(x_{i+1})$. It follows easily that

$$|\hat{H}_n(y) - H(y)| \le \max_i |\hat{H}_n(x_i) - H(x_i)| + \max_i |H(x_{i+1}) - H(x_i)|$$

for all $0 \le y \le M$ for all sufficiently large n w.p. 1. The first term on the right tends to zero by the theorem; the second term approaches zero since $\max_i |x_{i+1} - x_i| \to 0$ w.p. 1 and H is assumed to be continuous. Therefore by Polyá's Theorem, it suffices to show that

$$\hat{H}_n(\infty) \to H(\infty).$$

This follows easily from Theorem 3.1 since $x_n \to \infty$ w.p. 1 and

$$\hat{H}_n(\infty) - H(\infty) = \hat{H}_n(x_n) - H(x_n) + H(x_n) - H(\infty).$$

Theorem 3.2. *If F is strictly increasing on its support, F and H are continuous,*

$F(0) = 0$, $H(x) > 0$ for $x > 0$, and $H(0) = 0$, then

$$\sup_x |\hat{F}_n(x) - F(x)| \to 0 \quad \text{in probability as } n \to \infty. \tag{3.22}$$

The following lemmas are needed in the proof.

Lemma 3.4 (Bernstein). *If Z_i, $i \geq 1$, are independent and identically distributed random variables with zero mean and $-1 \leq Z_i \leq 1$, then*

$$P\left\{ \sup_{n \geq m} \bar{Z}_n > r \right\} \leq \exp\left\{ -\frac{m}{6} r^2 \right\}, \quad r \leq 1. \tag{3.23}$$

For a proof of this lemma see Serfling (1980), p. 95.

Lemma 3.5. (Helley). *If $P_n \Rightarrow P$, $g_n \to g$ uniformly where g is bounded and continuous a.e.(P), then*

$$\int g_n \, dP_n \to \int g \, dP \quad \text{as } n \to \infty. \tag{3.24}$$

This follows easily from Theorem 5.2, p. 31, of Billingsley (1968).

Proof of Theorem 3.2. It suffices to show convergence for fixed $x \geq 0$ by a simple variation on Polyá's Theorem. For $x \geq 0$ and $n \geq 1$, let

$$C_n(x) = \int_x^\infty \frac{1}{1 - e^{-\hat{H}_n(y)}} \, dF_n^\#(y). \tag{3.25}$$

Then $C = C_n(0)$ and $1 - \hat{F}_n(x) = C_n(x)/C_n(0)$, for all n, and x. So, it suffices to show that

$$C_n(x) \to \int_x^\infty \frac{1}{1 - e^{-H(y)}} \, dF^\#(y) = C(x)$$

say, for all x. For $x > 0$, this follows from Lemma 3.5 since $H(x) > 0$ for $x > 0$ and $1/(1 - e^{-\hat{H}_n}) \to 1/(1 - e^{-H})$ uniformly by Theorem 3.1. The delicate part is at $x = 0$. Let

$$\beta_n(\varepsilon) = \frac{1}{n} \sum_{i \, : \, x_i \leq \varepsilon} \frac{1}{1 - e^{-\hat{h}_i}}; \tag{3.26}$$

then $C = \beta_n(\varepsilon) + C_n(\varepsilon)$, for all $\varepsilon > 0$. Therefore it is enough to show that

$$\lim_{\varepsilon \to 0} \lim_n \beta_n(\varepsilon) = 0 \quad \text{in probability.} \tag{3.27}$$

Next, let $0 < c_n < \frac{1}{2}$, $n \geq 1$, and $m = m_n$, $n \geq 1$, be two sequences for which $c_n \to 0$;

$m \to \infty$; $m/nc_n \to 0$; and $ne^{-\eta mc_n^2} \to 0$ for all $\eta > 0$. For example if $c_n = n^{-1/5}$ and $m = \sqrt{n}$ for all $n \geq 1$, then the above conditions are satisfied. Now u has a negative derivative on $[0, \infty)$. So u^{-1} has a negative derivative on $(0, \frac{1}{2}]$. It follows that there is a $\Delta > 0$ for which $\frac{1}{2} - u(y) \geq \Delta y$ and $u^{-1}(\frac{1}{2} - y) \geq \Delta y$ for all sufficiently small y. Let

$$\delta_n = u^{-1}(\tfrac{1}{2} - c_n) \tag{3.28}$$

and

$$\beta_n'(\varepsilon) = \frac{1}{n} \sum_{i : x_i \leq \varepsilon, \, i > m} \frac{1}{1 - e^{-\hat{h}_i}}. \tag{3.29}$$

Then

$$|\beta_n - \beta_n'| \leq \frac{1}{n} \sum_{i \leq m} \frac{1}{1 - e^{-\hat{h}_i}} \leq \frac{m}{n\delta_n} \to 0.$$

Let

$$J = J(n, \varepsilon) = \{i \leq n : x_i \leq \varepsilon, \, H(x_i) \geq \delta_n\}, \tag{3.30}$$

and

$$\beta_n''(\varepsilon) = \frac{1}{n} \sum_{i \in J} \frac{1}{\hat{h}_{m+i}}. \tag{3.31}$$

Then

$$\beta_n'(\varepsilon) - \beta_n''(\varepsilon) \leq \frac{1}{n} \sum_{i : H(x_i) \leq \delta_n} \frac{1}{\hat{h}_{m+i}} \leq \frac{1}{n\delta_n} \#\{i : H(x_i) \leq \delta_n\},$$

and

$$E\{\text{RHS} \mid n\} = \frac{1}{\alpha\delta_n} \int_{\{H(x) \leq \delta_n\}} [1 - e^{-H(x)}] \, dF(x)$$

$$\leq \frac{1}{\alpha\delta_n} [1 - e^{-\delta_n}] F \circ H^{-1}(\delta_n) = o(1), \quad \text{as } n \to \infty,$$

by right continuity of F. So it is enough to show that $\lim_{\varepsilon \to 0} \lim_n \beta_n''(\varepsilon) = 0$. Let $\gamma_i = \frac{1}{2}(\frac{1}{2} - \theta_i)$, $i = 1, 2, \ldots, n$. Then $\gamma_i = \frac{1}{2}\{\frac{1}{2} - u[H(x_i)]\} \geq \frac{1}{2}c_n$, $\forall i \in J$. By definition of $\hat{\theta}_{m+i}$,

$$\hat{\theta}_{m+i} - \theta_i \leq \hat{\tilde{\theta}}_{m+i} - \theta_i \leq \max_{r \geq m} \frac{1}{r+1} \sum_{j=i}^{r+i} (t_j - \theta_j).$$

So,

$$P\{\hat{\theta}_{m+i} - \theta_i > \gamma_i, \, \exists i \in J \mid n, x_1, x_2, \ldots, x_n\}$$

$$\leq \sum_{i \in J} P\left\{\max_{r \geq m} \frac{1}{r+1} \sum_{j=i}^{r+i} (t_j - \theta_j) > \tfrac{1}{2}c_n \mid n, x_1, x_2, \ldots, x_n\right\}$$

$$\leq \sum_{i \in J} \frac{1}{A} \exp\left\{-\frac{m}{A}c_n^2\right\} \leq \frac{n}{A} \exp\left\{-\frac{m}{A}c_n^2\right\} \to 0 \quad \text{as } n \to \infty,$$

for some constant A. Finally, if $\hat{\theta}_{m+i} - \theta_i \leq \gamma_i$, for all $i \in J$, then

$$\hat{h}_{m+i} \geq u^{-1}(\theta_i + \gamma_i) = u^{-1}(\tfrac{1}{2} - \gamma_i) \geq u^{-1}[\tfrac{1}{2} - \Delta H(x_i)] \geq \Delta^2 H(x_i),$$

for all $i \in J$, for all sufficiently small ε. For such ε, let

$$\beta_n'''(\varepsilon) = \frac{1}{n} \sum_{i \in J} \frac{1}{\Delta^2 H(x_i)} \; . \tag{3.32}$$

Then

$$P\{\beta_n''(\varepsilon) \geq \beta_n'''(\varepsilon)\} \leq \frac{n}{A} \exp\left\{-\frac{m}{A} c_n^2\right\}$$

which is independent of ε and approaches zero as $n \to \infty$; and

$$E\{\beta_n'''(\varepsilon)\} \leq \int_0^\varepsilon \frac{1}{\Delta^2 H(x)} \, dF^\#(x) \leq \frac{1}{\Delta^2 \alpha} \int_0^\varepsilon \frac{1 - e^{-H}}{H} \, dF \to 0$$

as $\varepsilon \to 0$. It follows easily that $\beta(\varepsilon) \to 0$ in probability as $n \to \infty$ and $\varepsilon \to 0$ and therefore that $\hat{F}_n(x) \to F(x)$ in probability, for each fixed $x \geq 0$.

Corollary 3.2. *Under the conditions of Theorem 3.2,*

$$\hat{\alpha}_n \to \alpha \quad \text{in probability as } n \to \infty. \tag{3.33}$$

Proof. This follows directly from Theorem 3.2, since $\hat{\alpha}_n = 1/C_n(0)$.

Remark. It seems plausible that techniques developed by Groeneboom (1985) and P. Rao (1969) may be used to find the asymptotic distribution of $n^{1/3}(\hat{\theta}_j - \theta_j)$ as follows:

$$n^{1/3}(\hat{\theta}_j - \theta_j) \leq u \text{ iff } \hat{\theta}_j \leq \theta_j + \frac{u}{n^{1/3}} \text{ iff } U_n\left(\theta_j + \frac{u}{n^{1/3}}\right) \leq \frac{j}{n} \tag{3.34}$$

where the process U_n is defined by

$$U_n(a) = \sup\{t > 0 : \hat{K}_n(t) - at \text{ is maximal}\}. \tag{3.35}$$

This suggests using the empirical process J_n defined by

$$\{U_n(a) \leq t\} = \left\{\max_{s \leq t} J_n(s) > \max_{s > t} J_n(s)\right\}, \tag{3.36}$$

where, for fixed t,

$$J_n(s) = \hat{K}_n(s) - \hat{K}_n(t) - a(s - t). \tag{3.37}$$

It is reasonable to hope that the distribution of $J_n^*(u) = n^{2/3} J_n(t + u/n^{1/3})$ may be related to Brownian Motion and Brownian Bridges by using some results in Csörgő (1983). This suggests that the estimates converge at a rate of $n^{1/3}$.

all this possible. I am especially thankful to him for his guidance, his help, and his care throughout the research.

References

Barlow, R.E., D.J. Bartholomew, J.M. Bremner and H.D. Brunk (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

Barouch, E. and G.M. Kaufman (1975). Probabilistic modelling of oil and gas discovery. *Energy,* 133–152.

Bhattacharya, P.K., H. Chernoff and S.S. Yang (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11**, 505–514.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Breiman, L. (1968). *Probability.* Addison–Wesley, Reading, MA.

Csörgő, M. (1983). *Quantile Processes with Statistical Applications.* Regional Conference Series on Appl. Math. SIAM, Philadelphia, PA.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer,* Vol. II, 539–556.

Jackson, J.C. (1974). The analysis of quasar samples. *Mon. Not. R. Astr. Soc.* **166**, 281–295.

Kramer, M. (1983). Stopping a size dependent exploration process. Ph. D. Thesis, The University of Michigan.

Lee, P.J. and P.C.C. Wang (1986). Evaluation of petroleum resources from pool size distributions. In: D.D. Rice, Ed., *Oil and Gas Assessment – Methods and Applications, AAPG Studies in Geology No. 21.* Amer. Assoc. Pet. Geol., 33–42.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon. Not. R. Astr. Soc.* **155**, 95–118.

Marshall, A.W. (1970). Discussion of Barlow and van Zwet's papers. In: M.L. Puri, Ed., *Nonparametric Techniques in Statistical Inference.* Cambridge University Press, London, 175–176.

Nair, V.N. and P.C.C. Wang (1987). Statistical analysis of oil and gas discovery data. *AT&T Statistical Research Reports.* No. 44.

Nicoll, J.F. and I.E. Segal (1980). Nonparametric elimination of the observational cutoff bias. *Astron. Astrophys. L* **82**, 3–6.

Prakasa Rao, B.L.S. (1969). Estimating a unimodal density. *Sankhyā Ser. A* **31**, 23–26.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163–177.